



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **09321633 A**

(43) Date of publication of application: 12 . 12 . 97

(51) Int. Cl.

H03M 7/40
G06F 13/00
G06F 17/30

(21) Application number: **08136087**(22) Date of filing: **30 . 05 . 96**(71) Applicant: **NIPPON TELEGR & TELEPH
CORP <NTT>**

(72) Inventor: **KAJII TAKESHI
HAMANO TERUO
HASE MASAHICO
MURAKAMI MITSUO
SASAKI TSUTOMU
SONEHARA NOBORU**

(54) **INFORMATION ADDRESS CONVERSION
METHOD, INFORMATION ADDRESS
CONVERTER AND INFORMATION RETRIEVAL
SYSTEM**

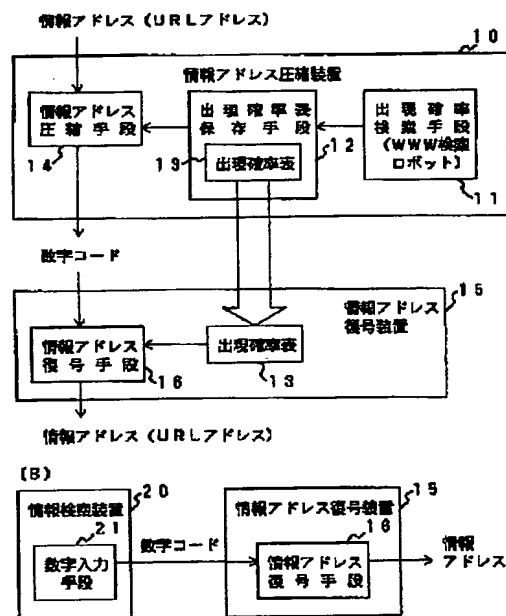
(57) Abstract:

PROBLEM TO BE SOLVED: To retrieve information of the world wide web(WWW) or the like on an internet with a simple device or a simple operation by designating an information address such as an URL address consisting of characters, numerals or symbol strings through the entry of a short numeral string.

SOLUTION: The incidence probability of characters, numerals or symbols incident in an information address is retrieved by a WWW retrieval robot method or the like and an obtained incidence probability table 13 is stored. Based on the incidence probability table 13, an information address being a character, numeral, symbol string is compression-coded into a numeral code by using a coding system such as Huffman coding or arithmetic coding. In the case of information retrieval, a numeral entry means 21 is used to enter a numeral code obtained through the conversion of the information address and an information address decoding means 16 uses the incidence probability table 13 to decode the code into

the original information address so as to retrieve the information.

COPYRIGHT: (C)1997,JPO
(A)



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-321633

(43) 公開日 平成9年(1997)12月12日

(51) Int. Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
H 0 3 M 7/40		8382-5K	H 0 3 M 7/40	
G 0 6 F 13/00	3 5 1		G 0 6 F 13/00	3 5 1 A
17/30			15/40	3 8 0 Z

審査請求 未請求 請求項の数 7 O L (全 7 頁)

(21) 出願番号 特願平8-136087

(22) 出願日 平成8年(1996)5月30日

(71) 出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72) 発明者 梶井 健

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(72) 発明者 坂野 昌夫

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(72) 発明者 長谷 雅彦

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

(74) 代理人 弁理士 小笠原 吉雄 (外1名)

最終頁に渡く

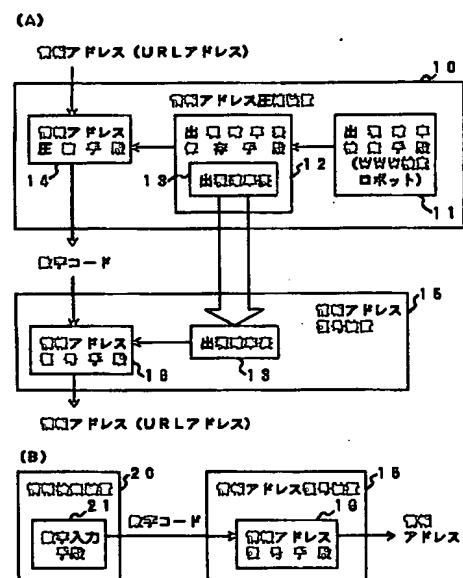
(54) 【発明の名称】 情報アドレス変換方法、情報アドレス変換装置および情報検索システム

(57) 【要約】

【課題】 文字や数字や記号列からなるURLアドレス等の情報アドレスを、短い数字列の入力により指定し、インターネット上のWWW等の情報検索を簡易な装置または簡易な操作によって行うことができるようにする。

【解決手段】 情報アドレスに出現する文字、数字または記号の出現確率をWWW検索ロボット法等により検索し、得られた出現確率表13を保存する。この出現確率表13をもとに、文字数字記号の列である情報アドレスをハフマン符号化や算術符号化等の符号化方式を用いて数字コードに圧縮符号化する。情報検索時には、情報アドレスを変換した数字コードを数字入力手段21により入力し、出現確率表13を用いて情報アドレス復号手段16により元の情報アドレスに復号し、検索を行う。

図1 情報アドレス変換装置の構成図



【特許請求の範囲】

【請求項1】 ネットワークを介して接続された情報蓄積装置が蓄積する所望の情報を検索するための情報の蓄積位置を示す情報アドレスを変換する方法であって、情報アドレスに出現する文字、数字または記号の出現確率を検索する出現確率検索過程と、検索した各文字数字記号の出現確率を保存しておく出現確率表保存過程と、該出現確率表保存過程で保存された出現確率をもとに、文字数字記号の列である情報アドレスを所定の圧縮符号化法を用いて数字コードに圧縮して変換する情報アドレス圧縮過程と、該数字コードとなった情報アドレスを前記圧縮符号化法に対応する所定の復号化法を用いて元の情報アドレスに復号する情報アドレス復号過程とを有することを特徴とする情報アドレス変換方法。

【請求項2】 ネットワークを介して接続された情報蓄積装置が蓄積する所望の情報を検索するための情報の蓄積位置を示す情報アドレスを変換する方法であって、情報アドレスに出現する文字、数字または記号の出現確率を検索する出現確率検索過程と、検索した各文字数字記号の出現確率をもとに、情報アドレスに出現する文字、数字もしくは記号またはこれらの組み合わせからなる各情報源記号を、所定の圧縮符号化法を用いて数字コードに変換し、各情報源記号と数字コードとの対応を示すコード表を作成するコード表作成過程と、該コード表作成過程で作成されたコード表をもとに、文字数字記号の列である情報アドレスを数字コードの列に圧縮して変換する情報アドレス圧縮過程と、該数字コードとなった情報アドレスを、前記出現確率または前記コード表を用いて前記圧縮符号化法に対応する所定の復号化法に基づく元の情報アドレスに復号する情報アドレス復号過程とを有することを特徴とする情報アドレス変換方法。

【請求項3】 前記情報アドレスは、インターネットにおけるURLアドレスであり、前記出現確率検索過程での出現確率検索法は、インターネット上にあるURLアドレスまたはキーワードの情報を自律的に検索するWWW検索ロボット法であることを特徴とする請求項1または請求項2に記載の情報アドレス変換方法。

【請求項4】 前記情報アドレス圧縮過程で用いる圧縮符号化法と、前記情報アドレス復号過程で用いる復号化法は、ハフマン符号化・復号化法または算術符号化・復号化法であることを特徴とする請求項1、請求項2または請求項3に記載の情報アドレス変換方法。

【請求項5】 ネットワークを介して接続された情報蓄積装置が蓄積する所望の情報を検索するための情報の蓄積位置を示す情報アドレスを変換する情報アドレス変換装置であって、情報アドレスに出現する各文字数字記号の出現確率をもとに、所定の圧縮符号化法により情報アドレスをあらかじめ圧縮符号化した数字コードを入力する数字入力手段と、情報アドレスの圧縮符号化に用いた各文字数字記号の出現確率をもとに、またはこの出現確

率を用いてあらかじめ作成された情報アドレスに出現する文字、数字もしくは記号またはこれらの組み合わせからなる各情報源記号とそれらを圧縮符号化した数字コードとの対応を示すコード表をもとに、前記数字入力手段によって入力した数字コードを、前記圧縮符号化法に対応する所定の復号化法に基づく元の情報アドレスに復号する情報アドレス復号手段とを備えることを特徴とする情報アドレス変換装置。

【請求項6】 ネットワークを介して接続された情報蓄積装置が蓄積する所望の情報を検索する情報検索システムにおいて、情報アドレスに出現する各文字数字記号の出現確率をもとに、所定の圧縮符号化法により情報アドレスをあらかじめ圧縮符号化した数字コードを入力する数字入力手段と、情報アドレスの圧縮符号化に用いた各文字数字記号の出現確率をもとに、またはこの出現確率を用いてあらかじめ作成された情報アドレスに出現する文字、数字もしくは記号またはこれらの組み合わせからなる各情報源記号とそれらを圧縮符号化した数字コードとの対応を示すコード表をもとに、前記数字入力手段によって入力した数字コードを、前記圧縮符号化法に対応する所定の復号化法に基づく元の情報アドレスに復号する情報アドレス復号手段と、ネットワークを介して、数字コードで入力された情報アドレスに該当する情報蓄積装置の情報を読み出す検索／読み出し手段と、該検索／読み出し手段により読み出された情報を表示する表示手段と、該表示手段において表示された情報とリンクした関連情報を選択操作する操作入力手段とを備えることを特徴とする情報検索システム。

【請求項7】 前記数字入力手段は、数字キーによる数字の入力手段またはバーコードによる数字の入力手段であることを特徴とする請求項6に記載の情報検索システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、インターネット等のネットワークを介して接続された情報蓄積装置から、情報検索装置を用いて所望の情報を検索し、提示する情報検索システムの技術分野に属し、特にインターネット上のホームページのアドレスであるURLアドレス等を、数字列で簡易に入力し指定することができるようにした情報アドレス変換方法、情報アドレス変換装置および情報検索システムに関する。

【0002】

【従来の技術】 インターネット上に構築されたWWW (World Wide Web) のように、ネットワークを介して接続された情報蓄積装置によって構成される分散型のデータベースでは、情報の蓄積位置を一意に表現するための情報アドレスの表記方法として、アルファベット、数字、記号等の文字数字記号を組み合わせたURLという表現が用いられている。検索したい情報の蓄積位置が既

知の場合、情報検索装置に検索したい情報のURLアドレスを入力して、情報を検索すればよい。

【0003】従来の情報検索装置では、アルファベット、数字、記号等の文字数字記号から構成されたURLを入力するためには、キーボードによる入力方法が一般的であった。このため、入力手段としてアルファベット等の文字入力キーがなく数字キーのみ、もしくはバーコード等の数字コード読み取りのみを有する情報検索装置では、URLアドレスの入力ができず、情報検索を行うことができなかった。

【0004】また、キーボード等の文字入力手段がある情報検索装置であっても、キーボードを操作した経験がないユーザには、キーボードは操作が煩雑であるという問題点があった。また、単に文字を表示しカーソルキー等で文字を決定する方法では入力に時間がかかってしまうという問題点があった。

【0005】ところで、WWWのようなハイパーテキスト構造の分散型データベースシステム上において、データベース上の情報およびその情報からハイパーテキスト構造により参照されている情報を再帰的に検索し、各情報のアドレス、ハイパーテキスト構造、検索用キーワード等のキーワード検索用データを収集するプログラムとして、WWW検索ロボットのようなものが知られている。なお、このWWW検索ロボット自体は、URLアドレスを数字に変換して入力するなどの、任意のURLアドレスの入力の簡易化を図るものではない。

【0006】参考文献：“The TkWWW Robot: Beyond Browsing”, Spetka, Scott, Proceedings of the Second International World-Wide Web Spetka, October 1994.

【0007】

【発明が解決しようとする課題】本発明の目的は、URLアドレスの入力手段として数字入力手段を有する情報検索装置において、数字入力によりURLアドレスの指定を行うためのURLアドレス指定方式を実現することである。

【0008】

【課題を解決するための手段】URLアドレスに代表されるような文字数字記号で表される情報アドレス（以下、URLアドレスとして説明する）を入力するための手段として、数字入力手段を用いるためには、URLアドレスを数字コードに変換する必要がある。しかし、単にURLアドレスの文字数字記号で表される文字コードを、数字コードに変換するだけでは、数字コードが長くなってしまふ。

【0009】そこで、平均的により短い数字列とするために、URLアドレスの文字列をハフマン符号化や算術符号化等の符号化方式を用いて、圧縮して数字列に変換する。一般的に、これらのデータ圧縮方法では、圧縮したい文字列をあらかじめ一回通読し、各文字の出現確率表を作成し圧縮データに添付する方法を用いる【参考文

献：情報と符号の理論、宮川ら、岩波書店、1982】。この方法では、十分に文字列が長い場合には、添付する出現確率表の大きさは問題とならないが、URLアドレスのように短い文字列の場合には、出現確率表自体の大きさが問題となる。

【0010】そこで、あらかじめWWW検索ロボット等によりURLアドレスを自律的に検索し、各文字もしくは文字列がURLアドレスに出現する確率を調べる。そして、得られた結果を出現確率表として用いて、URLアドレスを圧縮し数字コードに変換する。数字コードをURLアドレスに復号するときにも同じ出現確率表を用いればよい。

【0011】このことにより出現確率表を添付する必要がなくなる上、インターネットのアドレスに頻繁に使用される文字もしくは文字列に、より短い数字コードを割り当てることができ、単に文字を文字コードに置き換える場合と比較して、平均的にはより短い数字コードでURLアドレスを表現することが可能となる。

【0012】URLアドレスの入力を行うには、例えば図2に示すように、数字コードもしくはバーコード等、圧縮して数字列に変換されたURLアドレスを記録した紙等をあらかじめ用意しておく。圧縮して数字列に変換されたURLアドレスを作成するには、前記出現確率表を用いる。

【0013】操作者は、数字キーやバーコード読み取り手段等の数字入力手段により、所望の情報のURLアドレスに対応する数字コードの入力を行う。数字コードは、情報検索装置内もしくは他の情報蓄積装置上にある前記出現確率表をもとにURLアドレスに復号される。このことにより、所望した情報をインターネット等のデータベースから読み出してくることが可能となる。

【0014】出現確率表を直接用いる代わりに、出現確率表から作成した次のようなコード表を用いてもよい。情報アドレスに出現する文字、数字もしくは記号またはこれらの組み合わせからなる各情報源記号を、各文字数字記号の出現確率をもとにハフマン符号化法または算術符号化法などの圧縮符号化法を用いて数字コードに変換し、各情報源記号と数字コードとの対応を示すコード表を作成する。このコード表を記憶装置に記憶し、保存する。また、必要に応じて紙媒体に出力する。このコード表を用いれば、文字数字記号の列である情報アドレスを数字コードの列に圧縮して変換することができ、また、数字コードとなった情報アドレスを元の情報アドレスに復号することができる。

【0015】本発明により、URLアドレスの入力手段として数字入力手段を有する情報検索装置において、数字入力によりURLアドレスを指定し情報検索を行うことが可能となる。

【0016】

【発明の実施の形態】図1は、本発明に係る情報アドレ

ス変換説明図、図2は本発明を用いた情報アドレス変換の利用例を示す図、図3は情報アドレス変換の例を示す図である。

【0017】図1(A)に示すように、情報アドレス圧縮装置10は、出現確率検索手段11と、出現確率表13を保存する出現確率表保存手段12と、出現確率表13を用いて圧縮符号化を行う情報アドレス圧縮手段14からなる。また、情報アドレス復号装置15は、URLアドレスを変換した数字コードを、出現確率表13を用いて元のURLアドレスに変換する情報アドレス復号手段16を備える。

【0018】図1(B)に示すように、情報検索装置20は、インターネットに接続された情報蓄積装置の情報を検索する場合、あらかじめ情報アドレス圧縮装置10によってURLアドレスを圧縮符号化した数字コードを、数字入力手段21によって入力する。この数字コードを情報アドレス復号装置15の情報アドレス復号手段16に入力し、数字コードから元のURLアドレスを得て、検索を行う。このURLアドレスを表す数字コードは、以下に説明する方法によって短く圧縮したものにすることができる。

【0019】以下、図1ないし図3を参照し、URLアドレスを圧縮して数字列に変換する方法を具体的に説明する。図2に示す30は情報検索装置20に提供する情報を蓄積する情報蓄積装置、31は図1の情報アドレス圧縮装置10と情報アドレス復号装置15からなる情報アドレス圧縮・復号装置、32はLANやインターネット等のネットワークを表す。なお、情報アドレス圧縮装置10の部分は必ずしも情報蓄積装置30が持つ必要はない。

【0020】あらかじめ出現確率検索手段11で、WWW検索ロボット等の方法によりインターネット上にあるホームページのURLアドレスの情報を自律的に検索し、アルファベット、数字、記号等の各文字数字記号がURLアドレスに出現する確率を検索し、出現確率表13を出現確率表保存手段12に保存しておく。なお、WWW検索ロボットの詳細については、従来の技術の欄で述べた参考文献等から知られているので、ここでの詳しい説明は省略する。

【0021】情報アドレス圧縮装置10にURLアドレスが入力されると、出現確率表保存手段12に保存された出現確率表13をもとにして、ハフマン符号化や算術符号化等の圧縮方法によりURLアドレスを圧縮して数字列に変換し、URLアドレスを表す数字コードとして出力する。

【0022】一方、情報アドレス復号装置15にURLアドレスを表す数字コードが入力されると、出現確率表保存手段12に保存された出現確率表13をもとにして、ハフマン復号化や算術復号化等の方法により復号し、数字コードをURLアドレスに変換して出力する。

【0023】一般的にハフマン符号化や算術符号化等のデータ圧縮方法では、圧縮したい文字列をあらかじめ一回通読し、各文字の出現確率表を作成し圧縮データに添付する。これに対し本方法では、出現確率検索手段11により得られた出現確率表13をもとにデータの圧縮を行う。

【0024】ハフマン符号化・復号化法を用いた情報アドレス変換の例を図3に示す。変換対象となるURLアドレスが、例えば「http://S1S2S4S1S1S2」であったとする。先頭部分の「http://」は例えば既知であるとして、ここでは変換対象としない。図1(A)に示す出現確率検索手段11により、WWW検索ロボットを用いて検索した結果、情報源記号S1、S2、S3、S4のインターネット上のURLアドレスでの出現確率が、図3に示すようにそれぞれ3/7、2/7、1/7、1/7であったとする。各情報源記号に符号を対応させる手順は、以下のとおりである。

【0025】(1) まず、M個の情報源記号S1、S2、…、SMを出現確率の大きい順に並べる。

(2) 出現確率のもっとも小さい情報源記号2個をまとめ、これを一つの情報源記号におきかえ、その合成確率を新しい情報源記号の出現確率とする。この結果、記号数の一つ少ない新たな情報源記号の組が得られる。

【0026】(3) 再び確率の大きいものから順に並べ直す。

(4) 上記(2)、(3)の手続きを、最後に確率1の記号が残るまでくり返す。

(5) 以上のようにして符号の木をつくり、枝分かれて確率の大きい記号に0、小さい記号に1の符号を割り当て

る。

【0027】この符号に従って「http://S1S2S4S1S1S2」を変換すると「1000111100」となる。この方法により、URLアドレスに頻繁に現れる文字、もしくは文字列が多い場合には、より桁数の小さいビット数でURLアドレスを表すことができる。数字キーでの入力を簡単にするためには、2進数のコードを10進数に変換すればよい。

【0028】復号は、逆に符号木の根から出発して、符号化されたURLアドレスを一ビットずつ読み、0なら確率の大きい記号に、1なら確率の小さい記号のある枝に進む。葉に到着したら、その文字、もしくは文字列を出力し、根に戻ってくり返す。情報源記号には文字のみならず、例えば“www.”、“.jp/”等、URLアドレスに頻繁に現れる文字列を割り当ててもよい。

【0029】図3に示すコード表40は、URLアドレスに現れる各情報源記号S1、S2、…と、それらの出現確率をもとにハフマン符号化法によって圧縮符号化した数字コードとの対応表である。出現確率表13を用いる代わりに、このコード表40を作成して、URLアドレスの圧縮符号化・復号化に用いてもよい。

【0030】

【実施例】図4は、本発明の実施例によるシステム構成例を示す図である。図中、13は出現確率表、16は情報アドレス復号手段、20は情報検索装置、21は数字キーやバーコード読み取り手段等の数字入力手段、22は検索／読み出し手段、23は表示手段、24は操作入力手段、30は情報蓄積装置、32はインターネット等のネットワークを表す。

【0031】図4（A）のシステムでは、まず、紙等に印刷された数字コードに変換されたURLアドレスを、数字キー、バーコード読み取り等の数字入力手段21により入力する。数字コードとして入力されるURLアドレスは、図3で説明した情報アドレス変換方法によって圧縮して数字コードに変換されているものとする。

【0032】検索／読み出し手段22は、入力された数字コードをネットワーク32を介して情報アドレス復号手段16を有する情報蓄積装置30に送信する。情報アドレス復号手段16は、出現確率表保存手段12（図1）に保存されている出現確率表13（または図3に示すコード表40）をもとに、情報検索装置20から送信された数字コードを、URLアドレスに復号する。得られたURLアドレスを検索／読み出し手段22に返信して、検索／読み出し手段22でネットワーク32を介して、URLアドレスで指定された位置に蓄積された情報を読み出す。

【0033】URLアドレスを返信するのではなく、情報蓄積装置30内の検索／読み出し手段（図示省略）により、URLアドレスで指定された位置に蓄積された情報を読み出し、読み出された情報を情報検索装置20の検索／読み出し手段22に返信してもよい。得られた情報は表示手段23に出力される。表示手段23は、CRTや液晶ディスプレイ等のモニタ画面でもいいし、数字キーしかないファクシミリ装置のような紙への出力手段でもよい。検索した情報内に別情報へのリンクが含まれる場合には、リンクを操作入力手段24で選択し、情報の再検索を行う。

【0034】図4（B）に示すシステムの場合、情報蓄

積装置30が情報アドレス復号手段16を持つのではなく、情報検索装置20が持つ。この情報アドレス復号手段16が用いる出現確率表13は、例えばどこかの情報蓄積装置30がWWW検索ロボット等による出現確率検索手段11（図1）によって作成したものをあらかじめコピーして用意する。

【0035】数字入力手段21からURLアドレスを表す数字コードが入力されると、情報アドレス復号手段16は、出現確率表13によって元のURLアドレスに変換する。検索／読み出し手段22は、復元されたURLアドレスを用いて情報の検索を行う。

【0036】以上の構成により、数字入力によりURLアドレスの指定が可能となり、数字入力のみでインターネット上のWWW等の情報検索が可能となる。

【0037】

【発明の効果】本発明により、文字や数字や記号列からなるURLアドレス等の情報アドレスを、短い数字列の入力により指定し、インターネット上のWWW等の情報検索を簡易な装置または簡易な操作によって行うことができるようになる。

【図面の簡単な説明】

【図1】本発明に係る情報アドレス変換説明図である。

【図2】本発明を用いた情報アドレス変換の利用例を示す図である。

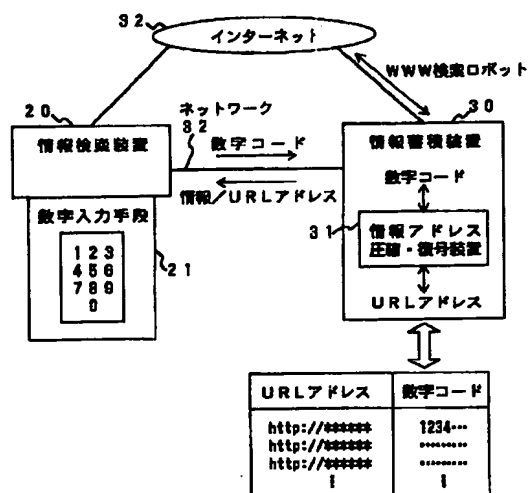
【図3】情報アドレス変換の例を示す図である。

【図4】本発明の実施例によるシステム構成例を示す図である。

【符号の説明】

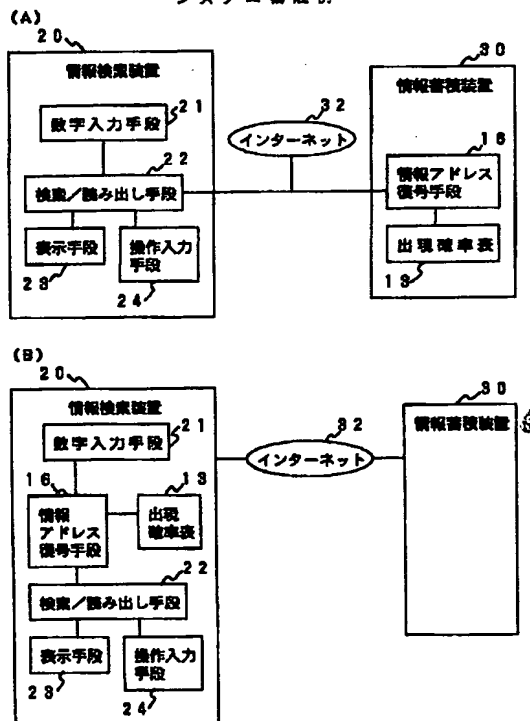
- 10 情報アドレス圧縮装置
- 11 出現確率検索手段
- 12 出現確率表保存手段
- 13 出現確率表
- 14 情報アドレス圧縮手段
- 15 情報アドレス復号装置
- 16 情報アドレス復号手段
- 20 情報検索装置
- 21 数字入力手段

【图2】



【图4】

システム構成例



フロントページの続き

(72)発明者 村上 満雄
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72)発明者 佐々木 努
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72)発明者 曾根原 登
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内